DOCUMENT RESUME

ED 071 747 PS 006 288

AUTHOR Potts, Marion
TITLE A Technique for Measuring Language Production in Three, Four & Five Year Olds.
NOTE 10p.; F-3454

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Data Analysis; *Language Development; Lower Class; Measurement Instruments; *Measurement Techniques; Middle Class; Morphemes; *Preschool Children; Speech; *Speech Evaluation; Syntax; *Tests; Verbs

ABSTRACT

The technique to be described here was developed to meet the need for a language production measure. It samples a variety of morphemic and syntactic patterns, and avoids the problems of imitation and of free speech analysis. The production test is administered in a sentence-completion format. A very brief incomplete story is told by the examiner in a form designed to elicit a particular target syntax. The test as developed contains 47 items covering 28 different structures, including present and past tense, auxiliary, possessives, negation, indirect object, conditional, relative clause, passive, and subjunctive. The final instrument was administered to 310 middle class children, and 163 lower class youngsters, all within the age range of 3/0 through 5/11. All children were individually tested by one of five white female examiners in rooms separate from the regular classroom. Data analyses were done by the three twelve-month groups of threes, fours, fives, as well as by six-month subgroups for the middle class. Data may indicate that acquisition of standard English patterns is not simply a bigger problem, but a different kind of problem for lower class children. (CK)

FILMED FROM BEST AVAILABLE COPY

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 071747

PS 006288

A Technique for Measuring Language Production In Three, Four & Five Year Olds

Marion Potts, Cornell University

The problems of studying language acquisition have been approached through techniques which probe comprehension, elicit imitation, or sample free speech. But when our interest lies in a child's productive control of particular syntactic structures, all of these measures are limiting. Comprehension level may differ from productive ability. The modelling aspect of imitation may result in a biased estimate of productive control. And free speech too often circumvents the structures of interest. A convenient method for eliciting syntactic structures could facilitate research into the rules by which children build sentences.

The technique to be described here was developed to meet the need for a language production measure. It samples a variety of morphemic and syntactic patterns, and avoids the problems of imitation and of free speech analysis. In addition, it taps structures which are seldom emitted in spontaneous speech, but which reflect command of critical language skills. We have tried to avoid contexts which trigger patterned responses, and have included items which require command of the relations among propositions. Because it was our hope to provide a measure which could be useful in early education as well as for acquisition research, the test focuses on the three, four and five year old age-range.

The production test is administered in a sentence-completion format. A very brief incomplete story is told by the examiner in a form designed to elicit a particular target syntax. As the story is told, a picture is shown which depicts what happened. Then when the story is stopped in mid-sentence, the child spontaneously blurts out the ending. For instance, one item which tests for a nominalization tells about some paint spilling on the floor, and shows a picture of a small girl wiping it up. The story ends with: "So Carol got a rag, and what she did next..." A correct response is one like "was wipe it up", requiring

use of the copula and deletion of the tense marker from the verb.

The final test was developed on the basis of data from a pilot test, which had been administered to 115 preschoolers. Piloting resulted in our dropping some items: those which were passed by almost all of the children, those which didn't differentiate between age levels, and those which were unreliable from test to retest. In addition, we saw the necessity of adding both ceiling items, and parallel items, to the final measure.

The test as developed contains 47 items covering 28 different structures, including present and past tense, auxiliary, possessives, negation, indirect object, conditional, relative clause, passive, and subjunctive. The final measure takes about ten minutes to administer.

As it is developed now, this performance technique is written only in standard English. It therefore reflects language competence only for children of standard English backgrounds. For this reason, the standardization group used to gather developmental data was comprised of middle class subjects. We did test lower class black and lower class white children, but for completely different reasons. Small samples of these groups were studied in an attempt to look at differing patterns in their language systems, and to identify structures which show ethnic and social class effects. This measure then, does not reflect level of language development for speakers of dialects. It taps their control only of standard English, not of their own well-developed systems.

The final instrument was administered to 310 middle class children, and 163 lower class youngsters, all within the age range of 3/0 through 5/11. Social class was determined through school records of parental occupation, with classes 6 and 7 on the Warner Scale defined as lower class membership. The samples were drawn from nursery schools, day care centers, Head Start and kindergarten programs in three cities of central New York State.

Due to the difficulty of finding a middle class black sample in this geographic area, the middle class group of 310 subjects was entirely white. Of the lower class children, 86 were black; 77 were white. Sex was evenly distributed within six-month age groups for each ethnic group and social class.

In addition to the Production technique, we planned to administer the Peabody Picture Vocabulary Test and the Bellugi-Klima Comprehension Tests of Grammatical Structures. The Peabody was chosen as a brief instrument which had already been standardized and which might reflect exposure to the language. We were aware of the cultural bias of this test for use with lower class children, and data were interpreted accordingly. The Comprehension Test uses three-dimensional materials, such as dolls, blocks, animals and trucks. Specific toys are set on a table for each item, and the child is asked to move the objects so as to show his understanding of the directions, e.g., "Show me - 'The dog is chased by the cat'."

Many of the structures tested by the Comprehension Test were also tapped by the Production Test. In addition, there were structures on each of the measures amenable to testing by one technique and not the other. The production measure had also dropped easier structures and added ceiling items on the basis of piloting. In spite of these differences, we were interested in seeing whether any relationship did exist between the two measures.

Procedure

All children were individually tested by one of five white female examiners in rooms separate from the regular classroom. It was not possible to find black examiners to test our black children, and this fact must be kept in mind in interpreting the data for that group. The Peabody was administered first, followed by the Production Test. One to three weeks later, half of all subjects were given the Comprehension Test and retested on the Production measure.

Production Test protocols of all subjects were taped during administration and transcribed later that day. All responses to each item were then listed and

categorized. For evaluating protocols, the final scoring system is a 0,1,2 point scale which is sufficiently clear-cut to be used by testers unsophisticated in language development. Questions for scoring were resolved by reference to the per cent of children at each developmental level who gave the response. For many items, the data indicated that one response was more advanced than another, even though the second was not incorrect. For instance, in a story about two children who had climbed high up into a tree, the most sophisticated response was a modal: "they'll get hurt" or "they might get hurt," and was given 2 points. "They're going to get hurt" was given proportionately less often by fives than by four and threes, and therefore received only one point. Deletion of the auxiliary on this item was scored zero, again substantiated by the data. Such developmental trends were checked for replication on our pilot sample of 115 subjects.

Our procedure then, involved leaning heavily on the shape of the age curves for each response which did not clearly conform to, or deviate from, standard English. We felt it necessary to investigate the validity of this procedure. If the response was language determined, this was an empirical way to make decisions. If however, responses were seriously affected by test orientation or socialization factors, our scoring procedure would contaminate the results. We planned to check on this issue in two ways: by studying the final test responses for replication of the trends of the pilot test, and by obtaining free speech samples to compare with responses on the final test.

RESULTS

It was our plan to examine the responses given for each structure in their relationships to age, social class, and ethnic group. Since we intended the measure to be an estimate of language development for middle class subjects, it was necessary to devise a rating system based on the data to provide a total performance index for each child. This "total score" was used in calculating the statistics given in Tables 1 and 2. As indicated previously, such an index for the lower class

samples reflects only conformity to, and deviation from, standard English as measured by this technique.

Data analyses were done by the three twelve-month groups of threes, fours, fives, as well as by six-month subgroups for the middle class. The scatterplot of 310 total raw scores for the middle class sample shows a clear linear trend, with variability decreasing with age. The plotting of means levels off at age five, where means for the six-month subgroups are not significantly different. At this level, middle class children begin to approach the ceiling of the test.

Regression lines and plotting of means also show linear trends for both lower class samples. However, the scatterplots reflect a marked difference in pattern between the middle class and both lower class groups. While the spread of the middle class children decreases with age, those of the lower class groups increase, resulting in clear triangular distributions. This pattern represents in part the wide variation in lower class language across the range from those who speak their dialects most pervasively, to those who are incorporating more elements of standard English.

Table 1 gives means and standard deviations for total raw scores on the Production Test for the middle class sample. Mean total scores show that the Production Test differentiates clearly among middle class twelve-month levels and shows less even developmental gains across the six-month groups.

Table 2 reports means and standard deviations separately for lower class blacks and whites. Breakdowns of the lower class samples are done among twelve-month levels only, due to the small n's in each age group. Standard English production scores reveal large differences, as would be expected, between both lower class groups on one hand and the middle class group on the other. There are much greater differences between middle class and lower class white children, than between lower class white and black children. Inspection of the lower class means however, shows that the latter two groups progressively separate from

PS 006288

each other with age. The lower class means for black and white five-year-olds are substantially farther apart than are those for the three-year-olds of these samples. The effect is just the opposite for the middle class white/lower class white comparison. However, the convergence of means with age in the latter groups is at least partly a function of the middle class children's approaching the ceiling of the test.

The different patterns of means across the three groups may be explained by differences in the developmental curves of individual structures. Some items show similar acquisition curves, though at different levels, for middle class and lower class whites. For structures influenced by ethnic background, curves for the lower class blacks are quite different. Responses such as deletion of the 3rd person singular morpheme, and of the copula, weigh heavily in the increasing differences with age in lower class black and white performance. Such responses, which are given by middle class and lower class th.ees, decrease with age in the two white groups, but stay at the same level or increase in our black children. The curves clearly demonstrate that these responses are not errors for black youngsters, but are part of a well-developed language system.

The similarity of some of the individual structure curves for middle and lower class whites needs further comment. It is only the acquisition curves which are approximately parallel; deviation curves of lower class and middle class whites are completely different from one another. This data may indicate that acquisition of standard English patterns is not simply a bigger problem, but a different kind of problem, for lower class children.

Reliability

As to reliability, three measures were obtained: test-retest correlations, individual item consistency, and internal consistency coefficients. We were especially interested in obtaining retest data to determine whether preschool

children would give the same types of responses over an interval of several weeks. Retest subjects comprised half of the original samples, and were randomly selected within the factors of sex, six-month age level, and classroom assignment. Total test-retest correlations were .92 for 156 middle class whites; .94 for 38 lower class whites; and .92 for 43 lower class blacks. Test retest r's within 6-month age groups range from .82 to .93; internal consistency r's by 6-month groups range from .84 to .88.

Contingency tables were set up to examine the proportion of subjects who responded consistently to each item on both final tests. Consistency in our framework meant that a child responded correctly, or incorrectly, on both administrations, even though the particular wording of his responses might vary. All items were responded to consistently by at least 70% of the subjects; three-quarters of the items by 80 to 90%.

Validity

Three types of validity data were examined: discrimination among age levels, concurrent validity of the three measures used, and relationship to free speech. Comparisons of mean performance at different ages has already been shown. We were also concerned with the extent to which individual items differentiated among threes, fours, and fives. Some items discriminate among all three age levels e.g., those for the passive and the relative clause structures. Elements such as subject pronoun and habitual present are developed early, and items tapping them differentiate between threes and fours. More difficult structures like indirect object and direct object, reciprocal, and subjunctive, showed greatest development between fours and fives. The addition of the negative to a structure which was also tested in its positive form consistently delayed development of the target response. Relative steepness of the curves does indicate where most gains are occurring, and clearly reflects developmental trends.

The second validity measure was the relationship among the three tests administered, as an index of concurrent validity. Intercorrelations of Peabody raw score and the Production Test were substantial: .67 for the total middle class group, .65 for lower class whites, and .55 for lower class blacks.

Correlations between the Production and the Comprehension Tests varied by social class but were significant for the total groups: .65 for the MC children, .39 for LC whites and .43 for LC blacks. Within 12 month age groups, all r's between production and comprehension were significant for MC children, and non-significant for LC children. This lack of significance probably represents the absence of developmental trends in the comprehension data for the small LC groups.

The third and most critical measure of validity is the relationship between free speech and performance on the production test. One hundred utterances of free speech data were gathered on each of 49 children who also took the test. Because our concern was the validity of the technique as a measure of standard English, all of these subjects were middle class children. For each free speech session, an individual child was taken to a private room, where an examiner encouraged him to play with a standardized set of toys and engaged him in conversation. All sessions were taped and immediately transcribed.

The protocols revealed that the children seldom generated obligatory contexts for structures they could not handle; they circumvented a difficult structure rather than use it incorrectly. Our analysis was therefore done by identifying all correct instances of each target structure in free speech, and by checking each matching test to determine whether that structure was used correctly in the test situation.

A major question we had was whether socializaton, attention, and testing factors might be accounting disproportionately for the missing of items by 3 year olds and other low scorers. Comparison of similarity of test performance with free speech usage shows a slight difference between threes and fives. There are

of course more instances of structures used by the older children: more structures are predicted to appear with development as the basis of the test. The average number of structures used correctly in both free speech and the test were 13.2 for the fives, 9.6 for the threes. Average number of structures used correctly in free speech but incorrectly on the test are 1.0 for fives, 1.2 for threes. Proportionately then, the fives do show fewer disconfirming utterances per number of structures used. The number of discrepancies for fives and threes however, is extremely slight.

The data was then examined for differences in children with high, middle, and low scores. The results are of the same pattern as those for the age comparison. They indicate that high scorers use most structures and show proportionately fewer inconsistencies. But differences in inconsistent responses throughout the score range are again very small.

Finally, the data was examined by structure. There were 340 instances of children correctly using in free speech those target structures which appeared on the test. Of these, in only 28 instances was there inconsistency between free speech usage and test performance. In other words, in only 8% of the cases did a child use a structure correctly in free speech which he used incorrectly on the test. We believe that this 92% consistency rate is supportive evidence of the validity of the technique.

Marion Potts
Cornell University

Table 1

Means and Standard Deviations of Total Raw Scores

on the Language Production Test: Middle Class Sample

Group	N	$\bar{X}$	SD
Early 3's: 3/0-3/5	41	42.49	16.37
Late 3's : 3/6-3/11	65	52.35	15.57
Total 3's	106	48.54	16.53
Early 4's: 4/0-4/5	45	55.29	15.28
Late 4's : 4/6-4/11	55	61.49	14.57
Total 4's	100	58.70	15.14
Early 5's: 5/0-5/5	49	73.74	12.85
Late 5's : 5/6-5/11	55	72.71	9.62
Total 5's	104	73.19	11.21
Total Sample	310	60.09	17.67

Table 2

Means and Standard Deviations of Total Raw Scores

on the Language Production Test: Lower Class Samples*

Group	N	$\bar{X}$	SD
Lower Class Whites			
Total 3's	26	17.08	11.15
Total 4's	24	34.63	19.75
Total 5's	27	50.15	20.71
Total Sample	77	34.14	22.28
Lower Class Blacks			
Total 3's	23	12.96	7.31
Total 4's	36	27.58	12.00
Total 5's	27	33.44	17.57
Total Sample	86	25.51	15.23

*This data reflects only degree of conformity to Standard English, not subjects' language development.